

**VOICE RECOGNITION EQUIPMENT**

**Publication number:** JP63104099

**Publication date:** 1988-05-09

**Inventor:** NODA SHIGETOSHI

**Applicant:** SONY CORP

**Classification:**

- **International:** **G10L15/10; G10L15/02; G10L15/00;** (IPC1-7):  
G10L7/08

- **European:**

**Application number:** JP19860250465 19861021

**Priority number(s):** JP19860250465 19861021

Report a data error here

Abstract not available for JP63104099

.....  
Data supplied from the **esp@cenet** database - Worldwide

⑬ 日本国特許庁(JP)

⑭ 特許出願公開

⑫ 公開特許公報(A) 昭63-104099

⑮ Int. Cl.<sup>4</sup>

識別記号

庁内整理番号

⑯ 公開 昭和63年(1988)5月9日

G 10 L 7/08

B-8221-5D

審査請求 未請求 発明の数 1 (全7頁)

⑰ 発明の名称 音声認識装置

⑱ 特 願 昭61-250465

⑲ 出 願 昭61(1986)10月21日

⑳ 発 明 者 納 田 重 利 東京都品川区北品川6丁目7番35号 ソニー株式会社内

㉑ 出 願 人 ソ ニ ー 株 式 会 社 東京都品川区北品川6丁目7番35号

㉒ 代 理 人 弁 理 士 杉 浦 正 知

明 細 書

1. 発明の名称

音声認識装置

2. 特許請求の範囲

入力音声信号から得られたスペクトルパターンをしきい値と比較することにより形成される二値の人力パターンと二値の標準パターンとのパターンマッチングを行う音声認識装置において、

上記入力パターン及び上記標準パターンの少なくとも一方の各フレームのデータに関してあるチャンネルの前及び後のチャンネルの夫々のデータの排他的論理和の値に基づき曖昧のチャンネルを判別し、

上記入力パターン及び標準パターンの対応するチャンネル間距離を求めると共に、上記判別結果に基づいて上記曖昧のチャンネル間の距離を減少させる補正を加えるようにしたことを特徴とする音声認識装置。

3. 発明の詳細な説明

(産業上の利用分野)

この発明は、例えば話者の音声単語単位で認識するのに適用される音声認識装置に関する。

(発明の概要)

この発明は、スペクトルデータの夫々を所定のしきい値で二値化してパターンマッチングを行う音声認識装置のパターンマッチング判定器において、1個のフレームの二値パターンのうちで音声の小さな変動に大きく影響されるしきい値近傍の値となるスペクトルデータを有していた曖昧的チャンネルを判定し、この判定出力に基づいて曖昧的チャンネルに関するチャンネル間距離を減少させる方向に補正する形でフレーム間距離を算出するようにしてマッチング精度を向上させ、音声の小さな変動による認識率の低下を防止するようにしたものである。

(従来の技術)

本願出願人により、先に提案されている特願昭59-106177号明細書に示される音声認識装置は、

音声入力部としてのマイクロホン、前処理回路、音響分析器、特徴データ抽出器、登録パターンメモリ及びパターンマッチング判定器等により構成されている。

この音声認識装置は、マイクロホンから入力される音声信号を前処理回路において、音声認識に必要とされる帯域に制限し、A/D変換器によりデジタル音声信号とし、このデジタル音声信号を音響分析器に供給する。

そして、音響分析器において、音声信号を周波数スペクトルに変換し、例えば対数軸上で一定間隔となるようにN個の間波数を代表値として周波数スペクトルを正規化して、フレーム周期毎にNチャンネルのスペクトルデータにより構成されるフレームデータの特徴データ抽出器に供給する。

特徴データ抽出器は、隣り合うフレームデータの距離を計算し、夾々のフレーム間距離の総和により、音声信号の始端フレームから終端フレームまでのN次元ベクトルの軌跡長を求め、最も語数が多く長い音声の場合に特徴を抽出するのに必要

な所定の分割数をもって軌跡長を等分割し、その分割点に対応したフレームデータのみを特徴データとして抽出して、話者の音声の発生速度変動に影響されることがないように時間軸を正規化し出力する。

この特徴データを登録時においては、登録パターンメモリに供給して登録特徴データブロック（標準パターン）として記憶し、認識時においては、入力音声信号を前述した処理により、入力特徴データブロック（入力パターン）とし、パターンマッチング判定器に供給する。そしてパターンマッチング判定器において、入力特徴データブロックと登録特徴データブロックとの間でパターンマッチングを行う。

パターンマッチング判定器は、登録特徴データブロックを構成するフレームデータと入力特徴データブロックを構成するフレームデータとの間でフレーム間距離を計算し、フレーム間距離の総和をマッチング距離とし、他の登録特徴データブロックに関しても同様にマッチング距離を算出して、

マッチング距離が最小で十分に距離に近いものと判断される登録特徴データブロックに対応する単語を認識結果として出力する。

しかし、従来の音声認識装置においては、音響分析器から出力されるフレームデータが特徴データ抽出器を介してそのまま登録特徴データブロックとして登録パターンメモリに記憶されるため、登録パターンメモリのメモリ量が膨大なものとなる問題点があった。これと共に、パターンマッチング時においても、データ量に応じてその計算処理時間が長くなる問題点があった。

このため、フレームデータを構成するスペクトルデータの々々を二値化し、登録パターンメモリの容量を低減させてマッチング処理時間を短縮化する音声認識装置（特開昭 60-166191号明細書）が本願出願人により提案されている。

この音声認識装置は、種々の原因により変動するスペクトルの傾向を補正するための傾向値を算出し、この傾向値に基づいてスペクトルの傾向を平坦化して話者の個人差や問題ノイズ等に影響さ

れることがないようにフレームデータを正規化した後にフレームデータを構成するスペクトルデータの々々を二値化処理し、得られた二値パターンに基づいてパターンマッチングを行うものである。

（発明が解決しようとする問題点）

しかし、前述した音声認識装置の二値化処理においては、適当に設定されたしきい値で以て各スペクトルデータが二値化され、例えばしきい値より大となるスペクトルデータが「1」とされ、しきい値より小となるスペクトルデータが「0」とされる。

例えば、第5図に示すチャンネル1～チャンネル12までの12個のスペクトルデータにより構成されるフレームデータが図中のしきい値で以て二値化された場合には、「0,1,1,1,1,0,0,1,0,1,1,0」の二値パターンとされる。しかし、しきい値近傍の値となるスペクトルデータは、やや乱雑な話し方や多少のノイズの混入によって変動し、二値化処理の結果として常に1（又は0）となる

とは限らず、入力時の状況により変化する曖昧なデータとなる。

つまり、音声特徴をより正確かつ適確に示すためには二値パターン中の「1」となるチャンネルがホルマント部と合致することが好ましいもので、第5図におけるチャンネル2、5のようにホルマントの両側となるチャンネルが「1」で示されることは、マッチング時における精度に悪影響を及ぼし好ましくない。また、第5図におけるチャンネル10、11のように、実際にはチャンネル10及び11との中間にホルマントが存在するにもかかわらず、近傍にずれた形でチャンネル10及び11の両者が「1」で示されることは、適確でない。

このように、しきい値近傍の値とされるスペクトルデータのチャンネルの二値データは曖昧であり、このような曖昧性を考慮せずに二値パターンに基づいてパターンマッチングする従来の音声認識装置は、正確でかつ適確なマッチングを行うことができず、音声の小変動に認識率が大きく影響

され易い問題点があった。

従って、この発明の目的は、二値化処理におけるしきい値近傍の曖昧のチャンネルを考慮した形で計算処理を行うことができ、マッチング精度が向上されると共に、認識率が向上された音声認識装置を提供することにある。

#### (問題点を解決するための手段)

この発明は、入力音声信号から得られたスペクトルパターンをしきい値と比較することにより形成される二値の入力パターンと二値の標準パターンとのパターンマッチングを行う音声認識装置において、入力パターン及び標準パターンの少なくとも一方の各フレームのデータに関してあるチャンネルの前及び後のチャンネルの夫々のデータの排他的論理和の値に基づき曖昧のチャンネルを判別し、入力パターン及び標準パターンの対応するチャンネル間距離を求めると共に、判別結果に基づいて曖昧のチャンネル間の距離を減少させる補正を加えるようにしたことを特徴とする音声認識

装置である。

#### (作用)

パターンマッチング判定器11において、登録特徴データブロックと入力特徴データブロックとの間で対応するフレーム $m$  ( $1 \leq m \leq M$ 、 $m$ は整数)のチャンネル $n$  ( $1 \leq n \leq N$ 、 $n$ は整数)に関して、そのチャンネル $n$ が音声の小変動に大きく影響される二値化処理におけるしきい値近傍のスペクトルデータを有していた曖昧のチャンネルかどうかチャンネル $n$ の前後に位置するチャンネル $(n-1)$ 及びチャンネル $(n+1)$ の二値データの排他的論理和に基づいて判定される。この判定出力に基づいて例えば1以下の補正係数 $W$ が曖昧のチャンネルに関するチャンネル間距離に乘せられ、チャンネル間距離を減少させる方向に補正する形で計算処理がなされてフレーム間距離が算出され、フレーム間距離に基づいて算出されたマッチング距離に基づいてマッチング判定がなされる。

#### (実施例)

##### a. 一実施例の全体の構成

以下、この発明の一実施例を図面を参照して説明する。第1図は、この発明の一実施例を示すもので、第1図において1で示されるのが音声入力部としてのマイクロホンを示している。

マイクロホン1からのアナログ音声信号がフィルタ2に供給される。フィルタ2は、例えばカットオフ周波数7.5kHzのローパスフィルタであり、音声信号がフィルタ2において、音声認識に必要とされる7.5kHz以下の帯域に制限され、この音声信号がアンプ3を介してA/D変換器4に供給される。

A/D変換器4は、例えばサンプリング周波数12.5kHzの8ビットA/D変換器であり、音声信号がA/D変換器4において、アナログ→デジタル変換されて8ビットのデジタル信号とされ、音響分析器5に供給される。

音響分析器5は、音声信号を同波数スペクトル

に変換して、例えばNチャンネルのスペクトルデータ列を発生するものである。音響分析器5において、音声信号が演算処理により周波数スペクトルに変換され、例えば対数軸上で一定間隔となるN個の周波数を代表値とするスペクトルデータ列が得られる。従って、音声信号がNチャンネルの離散的な周波数スペクトルの大きさによって表現される。そして、単位時間(フレーム周期)毎にNチャンネルのスペクトルデータ列が1つのフレームデータとして出力される。即ち、フレーム周期毎に音声信号がN次元ベクトルにより表現されるパラメータとして切り出され、スペクトル傾向正規化器6に供給される。

スペクトル傾向正規化器6において、順次供給されるフレームデータ毎にスペクトルデータの傾向正規化処理がなされる。例えば、各フレームデータを構成するスペクトルデータに関して傾向変動を補正する傾向値がチャンネル1から所定のチャンネルn( $1 \leq n \leq N$ , nは整数)までのスペクトルデータの平均値と、所定のチャンネルnか

ら最大チャンネルNまでのスペクトルデータの平均値との平均値に適当な係数が乗せられることにより求められる。この各チャンネルのスペクトルデータに関して求められた傾向値と対応するスペクトルデータとの間において減算がなされ、スペクトル傾向が平坦化され、話者の個人差及び周囲ノイズ等に影響されることがないようにスペクトル傾向が正規化される。全てのフレームに関して同様に傾向正規化処理がなされ、傾向正規化されたフレームデータが特徴データ抽出器7に供給される。

特徴データ抽出器7は、時間軸を正規化することにより時系列フレームデータを圧縮するものである。例えば、特徴データ抽出器7において、隣り合うフレームデータの各チャンネルに関してスペクトルデータの差の絶対値が夫々求められ、その総和が隣り合うフレームデータのフレーム間距離とされる。更に、フレーム間距離の総和が求められ、音声信号の始端フレームから終端フレームまでのN次元ベクトルの軌跡長が求められる。そ

して最も語数が多く長い音声の場合に特徴を抽出するのに必要な所定の分割数をもって軌跡長が等分割される。分割点の夫々に対応したフレームデータのみが抽出され、話者の音声の発生速度変動に影響されることがないように時間軸が正規化されて出力される。抽出されたフレームデータが二値化回路8に供給される。

二値化回路8において、適当な値に設定されたしきい値で、フレームデータを構成する8ビットのスペクトルデータの夫々が二値化される。例えば、スペクトルデータと適当な値に設定されたしきい値とが比較され、しきい値より大となる値のスペクトルデータが「1」とされ、しきい値より小となる値のスペクトルデータが「0」とされ、得られた1フレームに関する二値パターンがモード切替回路9に供給される。

この二値パターンが登録時においては、モード切替回路9を介して登録パターンメモリ10に供給され、例えば特徴データ抽出器7においてM個のフレームが抽出された場合には第2図に示すよ

うなデータブロックが登録特徴データブロックとして記憶される。認識時においては、入力音声信号が前述した処理を経て二値パターンとされ、この二値パターンがパターンマッチング判定器11に供給され、入力特徴データブロックとされる。入力特徴データブロックと比較の対象とされる全ての登録特徴データブロックとの間において、パターンマッチングが行われる。

b. この発明の一実施例におけるパターンマッチング判定器の説明

第3図は、パターンマッチング判定器11の一例を示し、第3図に示すように、フレーム距離計算回路12、曖昧判定回路13、マッチング距離計算回路14及び最小距離判定回路15によりパターンマッチング判定器11が構成される。

二値化回路8から入力特徴データブロックがフレーム距離計算回路12及び曖昧判定回路13に供給されると共に、登録パターンメモリ10から比較の対象とされる登録特徴データブロックがフレーム距離計算回路12及び曖昧判定回路13に

供給される。

フレーム距離計算回路12において、入力特徴データブロックと登録特徴データブロックとの対応するフレームのフレーム間距離が求められる。例えば、 $m$  ( $1 \leq m \leq M$ ,  $m$ は整数)番目のフレーム間距離 $D_m$ は、 $n$  ( $1 \leq n \leq N$ ,  $n$ は整数)をチャンネル番号とし、入力特徴データブロックの入力二値パターンを $S_{mn}$ とし、登録特徴データブロックの登録二値パターンを $R_{mn}$ とすると、次式により算出される。

$$D_m = \left\{ \sum_{n=1}^N |R_{mn} - S_{mn}|^P W \right\}^{1/P} \quad \dots (1)$$

尚、上記(1)式において( $P=1$ )の時はフレーム間距離 $D_m$ は、絶対値距離で算出され、( $P=2$ )の時はフレーム間距離 $D_m$ はユークリッド距離で算出される。

また、上記(1)式における $W$ は、補正係数であり、チャンネル $n$ が曖昧判定回路13の出力により曖昧的チャンネルと判定された時にのみ1以下の値とされ、曖昧的チャンネルと判定されない時には

1とされる。

曖昧判定回路13において、入力特徴データブロック及び登録特徴データブロックのフレーム $m$ におけるチャンネル $n$ の前後に位置するチャンネルの二値データの排他的論理和が次式に示すように求められる。

$$R_{m(n-1)} \oplus R_{m(n+1)} \quad \dots (2)$$

$$S_{m(n-1)} \oplus S_{m(n+1)} \quad \dots (3)$$

この上記(2)式及び(3)式により求められる二つの排他的論理和の少なくとも一方が「1」とされる時、そのチャンネルが曖昧的チャンネルと判定され、判定データがフレーム距離計算回路12に供給される。従って、フレーム距離計算回路12では、曖昧判定回路13からの判定データに基づいて曖昧的チャンネル以外では( $W=1$ )とされて計算処理され、あるチャンネルが曖昧的チャンネルと判定された時のみ $W$ が1以下の値とされて対応するチャンネル間における距離 $|R_{mn} - S_{mn}|^P$ に $W$ が乗せられる。二値化処理に際してはきい値近傍の値とされていたチャンネルに関するチャ

ンネル間距離を減少させる方向に補正が加えられた形でチャンネル間距離が累算されてフレーム間距離が算出される。フレーム距離計算回路12において順次計算処理されて得られる所定のフレームのフレーム間距離データがマッチング距離計算回路14に供給される。

マッチング距離計算回路14において、マッチング距離 $D$ が次式により算出される。

$$D = \sum_{m=1}^M D_m$$

つまり、順次フレーム間距離が累算され、最大フレーム $M$ までのフレーム間距離が累算されると、この累算値がマッチング距離とされ、マッチング距離データが最小距離判定回路15に供給される。同様に比較の対象とされる全ての登録特徴データブロックと入力特徴データブロックとの間においてマッチング距離が算出されてマッチング距離データが最小距離判定回路15に供給される。

最小距離判定回路15は、マッチング距離が最小で十分に距離に近いものと判断される登録特徴

データブロックに対応する単語を認識結果として出力する。

尚、この発明の一実施例においては、曖昧判定回路13において対応するフレームの全てのチャンネルに関して曖昧判定を行う場合について説明したが、入力二値パターン $S_{mn}$ または登録二値パターン $R_{mn}$ が「1」(若しくは0)の値でのみ判定処理を行うようにしても良い。

また、この発明の一実施例においては、フレーム距離計算回路12の計算処理においてチャンネル間距離 $|R_{mn} - S_{mn}|^P$ に補正係数 $W$ を乗ずる場合について説明したが、例えば( $P=1$ )の累加計算の時にはチャンネル間距離 $|R_{mn} - S_{mn}|$ が「1」となる時「 $K$ 」( $1$ 以上の適当な定数、例えば $K=2$ )とし、「0」となる時「0」と対応させ、曖昧的チャンネルに関して $W$ ( $K$ より小なる適当な定数、例えば $W=1$ )を対応させるようにしても良く、曖昧的チャンネルに関する距離が減少する方向に補正されるようにすれば良い。

更に、この発明の一実施例においては、一対一

に対応する直接的なマッチングによってマッチング距離を算出する場合について説明したが、例えばDPマッチング等の種々のマッチングサーチ手法における距離計算にこの発明を適用しても良い。

尚、この発明は、ハードワイヤードの構成に限らず、マイクロコンピュータ又はマイクロプログラム方式を用いてソフトウェアにより処理を行うようにしても良い。

#### (発明の効果)

この発明では、パターンマッチング判定器において、登録特徴データブロックと入力特徴データブロックとの間に対応するフレーム $m$  ( $1 \leq m \leq M$ 、 $m$ は整数)のチャンネル $n$  ( $1 \leq n \leq N$ 、 $n$ は整数)に関して、そのチャンネル $n$ が音声の小変動に大きく影響される二値化処理におけるしきい値近傍のスペクトルデータを有していた曖昧的チャンネルかどうかがチャンネル $n$ の前後に位置するチャンネル( $n-1$ )及びチャンネル( $n+1$ )の二値データの排他的論理和に基づいて判定

される。この判定出力に基づいて例えば1以下の補正係数 $W$ が曖昧的チャンネルに関するチャンネル間距離に乘ぜられ、チャンネル間距離を減少させる方向に補正する形で計算処理がなされてフレーム間距離が算出され、フレーム間距離に基づいて算出されたマッチング距離に基づいてマッチング判定がなされる。

例えば、第4図A及び第4図Bに示すチャンネル1～チャンネル12までの12個の二値データより構成される明らかに類似した2個のフレームがあるとする。前述した(1)式における $P$ を( $P=1$ )とし、補正係数 $W$ を( $W=0.5$ )とし、入力二値パターン $S_{1..}$ または登録二値パターン $R_{1..}$ が「1」の時のみ曖昧判定処理を行うものとしてフレーム間距離 $D$ 。が算出される場合には、第4図Aに示すフレームに関しては図中破線で囲んで示すようにチャンネル2、5、10、11の夫々が曖昧的チャンネルと判定されると共に、第4図Bに示すフレームに関しては図中破線で囲んで示すようにチャンネル3、5の夫々が曖昧的チャンネルと

判定される。そして、曖昧的チャンネルのチャンネル間距離に補正係数( $W=0.5$ )が乗せられた形でチャンネル間距離が累算されてフレーム間距離 $D$ 。が

$$D_{1..} = 0 + 1 \times 0.5 + 0 \times 0.5 + 0 + 0 \times 0.5 + 0 + 0 + 0 + 0 + 1 \times 0.5 + 0 \times 0.5 + 0$$

と算出され、 $D_{1..} = 1$ となる。これに対して従来の音声認識装置のパターンマッチング判定器のフレーム間距離計算処理を用いた場合(補正係数 $W$ を常に1として計算する)には、フレーム間距離 $D'_{1..}$ が

$$D'_{1..} = 0 + 1 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 1 + 0 + 0$$

と算出され、 $D'_{1..} = 2$ となる。従って、この発明の一実施例における計算処理により得られるフレーム間距離 $D$ 。の値は従来の計算処理により得られるフレーム間距離 $D'$ 。の値に比べて1/2の距離として算出され、第4図A及び第4図Bに示される2個のフレームがより類似していることを示す。

上述の例から理解されるように、この発明に依れば、フレーム間距離を算出する際に曖昧的チャンネルに関するチャンネル間距離が減少する方向に補正がなされてフレーム間距離が算出されるため、音声の小変動に大きく左右されるしきい値近傍のスペクトルデータを有していたチャンネルの二値データの影響が低減される。従って、二値化の利点であるメモリの容量の低減化及びマッチング処理の高速性を損なうことなく、音声の小変動による認識率の低下を防止することができ、高精度でかつ適確なパターンマッチングが可能となる。

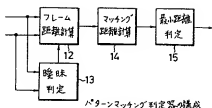
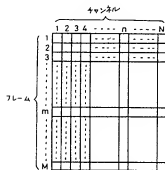
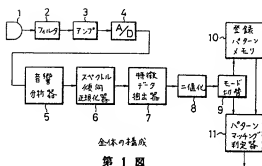
#### 4. 図面の簡単な説明

第1図はこの発明の一実施例の全体構成を示すブロック図、第2図はこの発明の一実施例における特徴データブロックのデータ構成の説明に用いる略線図、第3図はこの発明の一実施例におけるパターンマッチング判定器のブロック図、第4図はこの発明の一実施例におけるフレーム間距離計算の説明に用いる二値パターンの略線図、第5図は従来の音声認識装置の動作説明に用いる一例と

してのフレームデータである。

図面における主要な符号の説明

- 1 : マイクロホン、 5 : 音響分析器、  
6 : スペクトル傾向正規化器、 7 : 特徴データ  
抽出器、 8 : 二値化回路、 10 : 登録パター  
ンメモリ、 11 : パターンマッチング判定器、  
12 : フレーム距離計算回路、 13 : 曖昧判定  
回路、 14 : マッチング距離計算回路、  
15 : 最小距離判定回路。



A 0 ① 1 ① 0 0 1 0 ① 0

B 0 0 ① ① 0 0 1 0 0 1 0

第 4 図

